Language Runs on CT/HoTT: Compositional Semantics, Homotopical Identity, and a Testable Compliance Battery

Douglas C. Youvan

doug@youvan.com www.youvan.ai

September 14, 2025

This paper advances a simple claim with wide consequences: everyday human language already runs on the machinery of Category Theory (CT) and Homotopy Type Theory (HoTT). We show that ordinary composition of words and phrases behaves like categorical composition; that the grammar→meaning pipeline is a structure-preserving functor; that "meaning is use" is a direct Yoneda principle; and that ambiguity, coreference, and discourse coherence correspond to sums/products, pullbacks, and sheaf-style gluing. Crucially, sameness in language is intensional—better modeled by HoTT paths than by set-theoretic equality—so paraphrase, anaphora, and concept alignment are naturally higher-dimensional identity data. Building on these correspondences, we propose a compact, falsifiable test suite: metrics such as Paraphrase Homotopy Index (PHI), ua-witness tests for concept univalence, Pullback Coref Score (PCS), Beck–Chevalley defect for scope, Gluability index for discourse, and others. The point is not metaphor but measurement: each construct yields operational diagnostics that modern NLP systems can pass or fail. We argue this CT/HoTT view unifies classic insights across formal semantics, pragmatics, and distributional methods, offering clearer theory, better error localization, and concrete benchmarks for progress.

Keywords: category theory, homotopy type theory, compositional semantics, functoriality, naturality, yoneda lemma, products and sums, pullbacks, adjunctions, beck–chevalley, sheaf gluing, monads, univalence, function extensionality, transport, ∞-groupoids, paraphrase homotopy, coreference, quantifier scope, discourse coherence, metrics, benchmarks, nlp. A collaboration with GPT-5. CC4.0.

Outline:

1. Motivation and Thesis
   1.1 Why CT/HoTT for ordinary language
   1.2 From metaphor to measurement: our contributions and metrics
   1.3 Scope, audiences, and takeaways
2. Copy-Safe Math Conventions (Notation Box)
   2.1 ASCII operators and types (\circ, \Rightarrow, \Pi, \Sigma, =_A)
   2.2 Minimal term glossary (object, morphism, functor, naturality, path)
3. Language as a Category
   3.1 Compositionality as associativity: h \circ (g \circ f) = (h \circ g) \circ f
   3.2 Identity moves: id as neutral edits and empty contexts
   3.3 Well-typedness and selectional restrictions
4. Grammar → Meaning as a Functor
   4.1 Functoriality: F(id_A)=id_{F A}, F(g \circ f)=F(g) \circ F(f)
   4.2 Naturality as invariance across paraphrastic rewrites
   4.3 Diagnostic: Functor-consistency under benign grammar transforms
5. Meaning-as-Use via Yoneda
   5.1 Nat(Hom(A,-), F) ~= F(A) for lexical meaning profiles
   5.2 Typed probe batteries vs. untyped embeddings
   5.3 Metric: Yoneda-probe similarity
6. Products, Coproducts, and Case Analysis
   6.1 Coordination (products) and ambiguity (sums)
   6.2 Projections/injections as evidence extraction/disambiguation
   6.3 Dataset sketch and evaluation signals
7. Quantifier Scope, Adjunctions, and Beck–Chevalley
   7.1 Curry/uncurry: Hom(X \times A, B) ~= Hom(X, B^A) in relative clauses
   7.2 Scope as (co)reindexing; BC isomorphisms and failures
   7.3 Metric: BC-defect for scope ambiguities

8. Coreference as a Pullback
   8.1 Universal property for pronoun-antecedent alignment
   8.2 Feature functors and terminal cones
   8.3 Metric: Pullback Coref Score (PCS)

9. Discourse Coherence as Sheaf Gluing
   9.1 Local sections (sentences), overlaps (shared constraints)
   9.2 Global section existence and narrative breakdowns
   9.3 Metric: Gluability index

10. HoTT Identity for Linguistic Sameness
   10.1 Paths vs. equalities: p : s1 =_S s2 (paraphrase chains)
   10.2 Higher paths between paraphrases (2-cells)
   10.3 Metric: Paraphrase Homotopy Index (PHI)

11. Transport, Funext, and Univalence in Concepts
   11.1 transport^P(p,x) for substitution in context
   11.2 Function extensionality: contextwise equality ⇒ equality of expressions
   11.3 Univalence policy: Equiv(A,B) ~= (A =_U B) for concept collapse
   11.4 Metrics: Transport-invariance and ua-witness tests

12. Monads for Modality and Presupposition
   12.1 Effects and Kleisli composition (A -> T B)
   12.2 Belief/possibility/presupposition as typed side-effects
   12.3 Metric: Effect-leak audit

13. Polysemy as ∞-Groupoid Geometry
   13.1 Sense spaces with nontrivial higher cells
   13.2 Metric: Sense-nerve curvature and prediction difficulty

14. The CT/HoTT-Linguo Compliance Battery
   14.1 Metric definitions: PHI, ua-witness, PCS, BC-defect, Gluability, Yoneda-probes, Functor-consistency, Effect-audit, Sense-curvature, Transport-invariance, Curry-parity
   14.2 Aggregation and reporting

15. Mini Case Studies
    15.1 “Every student read some paper” (scope)
    15.2 Short coref paragraph (pullback)
    15.3 Paraphrase lattice with higher cells (PHI)
16. Related Work and Differentiation
    16.1 Categorial grammars, monadic semantics, sheaf models, distributional-categorical hybrids
    16.2 What is genuinely new here (integrated, testable, higher-path view)
17. Limitations, Risks, and Ethics
    17.1 Modeling choices and failure modes
    17.2 Interpretability gains vs. overformalization risks
18. Outlook and Future Work
    18.1 Benchmarks and open problems
    18.2 Toward CT/HoTT-native NLU systems
19. References

1. Motivation and Thesis

We claim that ordinary human language already runs on Category Theory (CT) and Homotopy Type Theory (HoTT). Composition of phrases behaves like morphism composition; the grammar→meaning pipeline is a structure-preserving functor; ambiguity/coreference/coherence are limits, coproducts, and gluing conditions; and linguistic "sameness" is intensional—better modeled as paths (and higher paths) than as plain equality. The goal is not metaphor but *measurement*: we turn these correspondences into falsifiable metrics that any NLU system can pass or fail.

1.1 Why CT/HoTT for ordinary language

- **Compositionality as structure, not slogan.** Language builds big meanings from small ones. CT says how: h \circ (g \circ f) = (h \circ g) \circ f (associativity) and id is neutral. This matches stable truth-conditions under benign rebracketings and edits.
- **Universal properties explain "why this, not that."** Products/sums, pullbacks, and adjunctions *characterize* constructions by what they do, not how they look. This mirrors coordination ("and"), disjunction/ambiguity ("or"), pronoun resolution, and scope.
- **Identity is path-like.** Paraphrase, anaphora, and concept alignment are *ways of being the same*, not mere equality. HoTT provides p : a =_A b, higher paths between paths, and transport for substitution in context.
- **Economy and modularity.** CT/HoTT separates *plumbing* (composition/typing) from *content* (lexical denotations), yielding clearer error localization and reusable proofs of invariance.
- **Bridging formal and data-driven.** Yoneda operationalizes "meaning is use"; typed probe batteries connect symbolic structure to measurable distributional behavior.

1.2 From metaphor to measurement: our contributions and metrics

We package a compact, falsifiable *CT/HoTT-Linguo compliance battery*:

- **Paraphrase Homotopy Index (PHI).** Distinct paraphrase routes as distinct 2-paths; counts nontrivial cycles in the paraphrase graph.
- **ua-witness test (Univalence).** Collapse concepts A,B only when we have learned Equiv(A,B) with round-trip witnesses: e^{-1} \circ e \simeq id_A, e \circ e^{-1} \simeq id_B.
- **Pullback Coref Score (PCS).** Coreference by solving feature equalities via a terminal cone (pullback); score = fraction of gold links recovered by the universal property.
- **Beck–Chevalley defect (scope).** Quantifies failure to commute reindexing with \Sigma/\Pi in quantifier-scope ambiguities.
- **Gluability index (discourse).** Measures whether local sentence constraints glue to a global section without contradictions.
- **Yoneda-probe similarity.** Typed context probes implementing Nat(Hom(A,-),F) \cong F(A); similarity predicts synonymy under controlled contexts.
- **Functor-consistency.** Tests F(id_A)=id_{F A} and F(g \circ f)=F(g) \circ F(f) by benign grammar rewrites that should preserve denotation.
- **Effect-leak audit (monads).** For modality/presupposition modeled as (A -> T B), detect unintended side-effects under composition.
- **Sense-nerve curvature.** Treat polysemy as an \infty-groupoid; quantify higher-cell structure among senses.
- **Transport-invariance.** For justified co-reference p : a = b, check downstream predictions unchanged after transport^P(p, x).
- **Curry-parity.** Verify Hom(X \times A, B) \cong Hom(X, B^A) on relative-clause variants.

Each metric pairs a *law* (copy-safe math) with a *behavioral signature* (dataset transform + pass/fail criterion).

1.3 Scope, audiences, and takeaways

- **Scope.** We target sentence- to paragraph-level semantics and discourse coherence. We use copy-safe ASCII math (\circ, \Pi, \Sigma, =_A, \simeq, id_A, ua) and avoid dependence on any single grammar formalism. The account is compatible with categorial grammars, typed lambda semantics, and distributional models.
- **Audiences.**
    - *Theorists:* a unifying CT/HoTT lens with universal properties explaining core linguistic phenomena.
    - *NLP practitioners:* a small, testable compliance battery to diagnose models beyond aggregate accuracy.
    - *Cognitive scientists:* operational proxies for compositionality, identity-of-meaning, and discourse integration.
- **Takeaways.**

1. Language exhibits categorical composition and universal-property characterizations.
2. Linguistic identity is homotopical: paths/higher paths matter.
3. These insights yield concrete, falsifiable metrics that can guide model design and evaluation.

2. Copy-Safe Math Conventions (Notation Box)

This paper uses **ASCII-safe, LaTeX-style inline math** (no Unicode symbols). All equations appear in backticks. We keep types and categories explicit and avoid overloaded notation.

2.1 ASCII operators and types we use

- **Types & arrows.**
  f : A -> B (a typed arrow); composition g \circ f : A -> C; identity id_A : A -> A.
  Products and sums: A \times B, A + B. Exponentials: B^A.
  Terminal/initial: 1, 0.
  Hom-sets: Hom_C(A,B) (often Hom(A,B) when C is clear).

- **Functors & naturals.**
  Functor: F : C -> D, action on arrows: F(f).
  Natural transformation: \alpha : F \Rightarrow G, components \alpha_A : F A -> G A.
  Functor laws: F(id_A)=id_{F A}, F(g \circ f)=F(g) \circ F(f).
  Naturality: G(f) \circ \alpha_A = \alpha_B \circ F(f).

- **Universal properties.**
  Products: projections \pi_1 : A \times B -> A, \pi_2 : A \times B -> B; pairing \langle f,g \rangle : X -> A \times B.
  Coproducts: injections inl : A -> A+B, inr : B -> A+B; case analysis [f,g] : A+B -> C.
  Pullback square: f \circ p1 = g \circ p2; universal m with p1 \circ m = u, p2 \circ m = v.
  Adjunction: F \dashv G with unit \eta : id \Rightarrow G F and counit \varepsilon : F G \Rightarrow id; triangle laws \varepsilon_{F A} \circ F(\eta_A) = id_{F A}, G(\varepsilon_B) \circ \eta_{G B} = id_{G B}.
  Exponential adjunction: Hom(X \times A, B) \cong Hom(X, B^A).

- **HoTT identity & transport.**
  Path/identity type: p : a =_A b. Reflexivity: refl_a : a =_A a.

Path induction (J): eliminates equalities back to refl.
Transport in family P : A -> Type: transport^P(p,x) : P(b) and transport^P(refl_a,x)=x.
Function extensionality:
funext : (forall x:A, f(x) =_B g(x)) -> (f =_{A->B} g).
Univalence (copy-safe): ua : Equiv(A,B) \simeq (A =_U B) with inverse ua^{-1}.
Action on paths: ap_f(p) : f(a) = f(b), with ap_f(refl_a)=refl_{f(a)}.

- **Equality/readability conventions.**
  We distinguish **definitional** from **propositional** equalities only in prose; inline we use = for equations of arrows/types and =_A for identity types. We write isomorphism/adjointness with \cong, \simeq, \dashv (all as plain ASCII strings).

- **Effects (monads).**
  Monad T with unit \eta : A -> T A and bind >>=. Kleisli arrows: A -> T B.
  We use this for modality/presupposition pipelines.

2.2 Minimal term glossary (plain language anchors)

- **Object/Type (A,B):** a category's "thing" or a HoTT type (e.g., entities, propositions, contexts).
- **Arrow/Morphism (f : A -> B):** a well-typed meaning-building step.
- **Composition (g \circ f):** do f then g; associative with identities neutral.
- **Functor (F):** grammar-to-meaning map that preserves structure.
- **Natural transformation (\alpha : F \Rightarrow G):** a coherent relabeling between functors.
- **Product/Coproduct:** pairing/branching; coordination vs. ambiguity.
- **Pullback:** most general way to keep two views consistent (coreference).
- **Adjunction:** "best-match" correspondence (scope, raising); triangle laws ensure no extra junk.

- **Exponential (B^A):** a “function object”; currying/relative clause behavior.
- **Identity path (a =_A b):** a *way* two things are the same (paraphrase/anaphora).
- **Transport:** carry data through a proven sameness.
- **funext / ua:** pointwise-same implies same function; equivalence of types equals equality of types (policy for concept collapse).
- **Sheaf gluing:** local sentence constraints that agree on overlaps assemble into a coherent discourse.

3. Language as a Category

We model a small, typed category Lang in which objects are linguistic/semantic types (e.g., NP, VP, S, Ent, Prop, Ctx) and arrows are well-typed meaning-building steps (lexical denotations, modifiers, combinators). Composition is phrase assembly; identities are neutral edits or empty contexts.

3.1 Compositionality as associativity

Associativity states that rebracketing benign combinations preserves denotation.

- Law: h \circ (g \circ f) = (h \circ g) \circ f.
- Typed toy: let tall : N -> N, very : (N->N) -> (N->N) (adjective intensifier), building : 1 -> N, say : N -> S (predicate “is …”). Then
- say \circ (very(tall) \circ building) = (say \circ very(tall)) \circ building .

Reading: whether we first make a “very-tall” concept then assert it of building, or first prepare the predicate then apply to building, truth-conditions match.

- Benign rebracketings we admit: adjunct/degree stacking, certain PP attachments under explicit isomorphisms, and local compositional plumbing where heads/arguments stay fixed.

3.2 Identity moves and neutral contexts

- Law: id_A \circ f = f and g \circ id_A = g.
- Linguistic reading: neutral edits (e.g., adding/removing vacuous parentheses in derivations, inserting an empty context, or performing a no-op formatting change) do not alter meaning. For a constituent of type A, the no-op arrow id_A : A -> A models “do nothing to A”.
- Example:
- id_S \circ (say \circ very(tall) \circ building) = say \circ very(tall) \circ building .

3.3 Well-typedness and selectional restrictions

Typing blocks category errors.

- If bark : Anim -> VP and the_dog : 1 -> Anim then bark \circ the_dog : 1 -> VP is well-typed.
- If idea : 1 -> Abstract and there is no inclusion Abstract -> Anim, the composite bark \circ idea is ill-typed (no arrow). Selectional restrictions are type constraints that regulate which compositions exist in Lang.
- Practical diagnostic: a parser/semanticist consistent with this view should reject (or repair via coercion maps) derivations that lack appropriate arrows.

---

4. Grammar → Meaning as a Functor

We posit a structure-preserving map F : Syn -> Sem from a syntactic category to a semantic category (e.g., from derivations to typed lambda terms or model-theoretic denotations).

4.1 Functoriality

- Laws:
- F(id_A) = id_{F A}
- F(g \circ f) = F(g) \circ F(f)

Reading: interpreting a neutral syntactic move yields a neutral semantic move; interpreting a composite derivation equals composing the interpretations.

- Example:
- F(say \circ very(tall) \circ building) = F(say) \circ F(very(tall)) \circ F(building) .

4.2 Naturality as invariance across paraphrastic rewrites

Between two semantic functors F,G : Syn -> Sem (e.g., truth-conditional vs. distributional semantics), a natural transformation \alpha : F \Rightarrow G has components \alpha_A : F A -> G A satisfying

G(f) \circ \alpha_A = \alpha_B \circ F(f) .

Reading: relabeling meanings commutes with interpreting derivations—paraphrase-preserving rewrites align across semantics.

4.3 Diagnostic: Functor-consistency under benign transforms

We define a **Functor-consistency** test suite:

- Construct pairs (d, d') where d' is obtained from d by a benign syntactic transform that theory predicts should not change denotation (e.g., rebracketing under associativity, trivial adjunct reassociation with explicit isomorphisms, insertion/removal of neutral format tokens).
- Require:
- F(d) = F(d') (semantic equality or specified isomorphism)
- Score: fraction of pairs passing; localize failures to either F(id)=id or F(\circ)=\circ violations.

5. Meaning-as-Use via Yoneda

We operationalize “meaning is use” by typed context probes and the Yoneda lemma.

5.1 Yoneda principle for lexical meaning profiles

For a semantic functor F : C -> Set and an object A (the lexical type of a word), Yoneda states:

Nat(Hom_C(A,-), F) \cong F(A) .

Reading: natural ways to map A-shaped contexts into meanings correspond one-for-one to the word’s denotation at A.

Interpretation: The behavior of a word across *all typed contexts* determines (and is determined by) its meaning. Practically, we approximate “all contexts” with a curated, finite, typed probe set.

5.2 Typed probe batteries vs. untyped embeddings

Design a **Yoneda-probe battery**:

- Choose typed context templates {c_i} with c_i \in Hom(A, X_i) (e.g., for A=N, templates like “_ is heavy”, “count the _”, “define the _” that are syntactically/semantically valid for nouns).
- For a word w of type A, compute probe responses {F(c_i)(w)} (truth-conditions, entailment signatures, or task-specific outputs).
- Define similarity between words w1,w2 by comparing their response vectors under the same {c_i}. This yields a *typed* alternative to raw distributional similarity.

5.3 Metric: Yoneda-probe similarity

Given probe responses embedded into a common representation space, define

YonedaSim(w1, w2) = sim({F(c_i)(w1)}_i , {F(c_i)(w2)}_i)

with sim a metric that respects type labels (e.g., cosine within each X_i, then aggregated). Hypothesis: true synonyms at type A maximize YonedaSim, while mere co-occurrence neighbors without typed interchangeability do not.

Dataset sketch: For each part-of-speech/type A, assemble minimal pairs (synonym, near-synonym, unrelated). Evaluate ranking by YonedaSim, report AUC and calibrated thresholds for synonymy detection.

6. Products, Coproducts, and Case Analysis

Products capture **coordination** ("and"); coproducts capture **branching/ambiguity** ("or", polysemy cases before disambiguation). Case analysis resolves sums.

6.1 Coordination (products) and ambiguity (sums)

- **Products (A \times B).** Projections and pairing:
- \pi_1 : A \times B -> A ,\quad \pi_2 : A \times B -> B ,
- \langle f , g \rangle : X -> A \times B

Laws:

\pi_1 \circ \langle f , g \rangle = f ,\quad \pi_2 \circ \langle f , g \rangle = g ,

\langle \pi_1 \circ h , \pi_2 \circ h \rangle = h .

Linguistic reading: "Alice and Bob left" factors to each leaving (\pi_1, \pi_2); any arrow into the coordinated meaning is determined by the two components.

- **Coproducts (A + B).** Injections and case analysis:
- inl : A -> A+B ,\quad inr : B -> A+B ,\quad [f,g] : A+B -> C

Laws:

[f,g] \circ inl = f ,\quad [f,g] \circ inr = g ,

and [f,g] is unique with those equations. Reading: a word with two compatible senses is an A+B; resolving sense is [f,g].

6.2 Projections/injections as evidence extraction/disambiguation

- **Evidence extraction.** From X -> A \times B, we recover each conjunct by composing with projections; this models entailments from a coordination.
- **Disambiguation.** From A+B, downstream constraints select a branch via [f,g]. Example: “bank” as Finance + River. In “the bank approved the loan,” the VP expects Finance, yielding [f,g] that kills the river branch.

6.3 Dataset sketch and evaluation signals

- **Coordination test.** Build pairs differing only by harmless rebracketing of conjunctions; require
- F(\text{A and (B and C)}) \cong F(\text{(A and B) and C})

Score: product-associativity accuracy.

- **Ambiguity test.** Provide sentences where a single token is an A+B and downstream contexts force A or B. Evaluate success of [f,g] selection; report per-branch accuracy and calibration.
- **Cross-task signal.** Projections (\pi_i) predict entailment patterns (e.g., from “Alice and Bob left” infer “Alice left”). Measure precision/recall of projection-derived entailments.

---

7. Quantifier Scope, Adjunctions, and Beck–Chevalley

Scope phenomena align with currying/adjunctions and (co)reindexing across pullback squares. Failures surface as **Beck–Chevalley** (BC) defects.

7.1 Curry/uncurry in relative clauses

Exponential adjunction:

Hom(X \times A, B) \cong Hom(X, B^A)

Reading: a predicate expecting (X \times A) is equivalent to a function that, given X, returns a predicate over A. Relative clauses often implement this movement:

“books [that Alice read]” corresponds to currying a 2-place relation into a unary modifier on books.

7.2 Scope as (co)reindexing; BC isomorphisms and failures

Represent **quantifiers** with dependent sums/products:

\Sigma_x : A . P(x) (existential) ,\quad \Pi_x : A . P(x) (universal)

A typical ambiguity:

Every student read some paper.

Two readings:

\Pi_{s:Student} \Sigma_{p:Paper} Read(s,p)

\Sigma_{p:Paper} \Pi_{s:Student} Read(s,p)

Interpret scopal rearrangement as commuting reindexing with \Sigma/\Pi along a square. When the square is **cartesian**, a BC isomorphism justifies commutation; otherwise, a **BC defect** explains preference or impossibility.

7.3 Metric: Beck–Chevalley defect (BC-defect)

- **Construction.** For minimal pairs that differ by scope (\Pi\Sigma vs \Sigma\Pi), build the associated square and label it cartesian or non-cartesian based on lexical/semantic constraints (e.g., specificity, islands).
- **Score.** A model’s logit margin favoring the BC-permitted reading minus the margin for the BC-blocked reading. Aggregate per condition:
- BCdefect = mean(margin_{cartesian} - margin_{noncartesian})
- **Prediction.** Robust scope handling yields large positive BCdefect; pathological scope errors show small or negative values. Localize failures to reindexing modules.

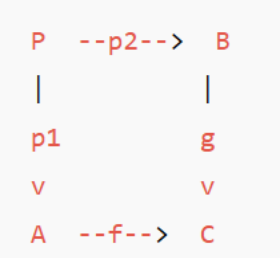
---

8. Coreference as a Pullback

Pronoun–antecedent linking is choosing a **terminal cone** equalizing constraints (number, gender, animacy, recency, discourse role).

8.1 Universal property for pronoun alignment

Let f : A -> C encode anaphoric **features** of candidate antecedents (A = set/type of mentions), and g : B -> C encode **requirements** of the pronoun (B often 1 with features). A **pullback**



satisfies f \circ p1 = g \circ p2 and is terminal with this property. Reading: P is the most general way to keep antecedent-features equal to pronoun-requirements.

8.2 Feature functors and terminal cones

Model feature extraction as a functor Feat : Disc -> C from a discrete category of mentions. Candidate choice is the unique mediating arrow m : 1 -> P (or a distribution over its elements) satisfying the pullback constraints. Competing constraints (agreement vs. salience) appear as additional legs commuting into C.

8.3 Metric: Pullback Coref Score (PCS)

- **Construction.** Implement a solver that forms P from gold or predicted features and selects the terminal solution; compare with gold coref links.
- **Score.** PCS = (# gold links recovered by pullback) / (total gold links). Report also *violation rate* where no solution exists due to inconsistent features—useful for debugging feature extractors.

---

9. Discourse Coherence as Sheaf Gluing

Coherent discourse is the existence of a **global section** consistent with all **local sections** (sentences/clauses) that agree on overlaps (shared entities, times, topics).

9.1 Local sections and overlaps

Let a discourse domain be covered by patches {U_i} (sentence-local contexts). Each sentence contributes a section s_i over U_i. On overlaps U_i \cap U_j, the restrictions must agree:

s_i |_{U_i \cap U_j} = s_j |_{U_i \cap U_j}

If such compatible families exist for a cover, a sheaf yields a glued global section s over \cup_i U_i.

9.2 Global section existence and narrative breakdowns

Breakdowns (topic drift, tense inconsistencies, presupposition failures) appear as **incompatibilities on overlaps**; no global s exists. Repairs correspond to modifying some s_i or refining the cover to restore agreement.

9.3 Metric: Gluability index

- **Procedure.** Encode each sentence's constraints into a typed store; compute pairwise overlap constraints; attempt a constructive gluing (e.g., via constraint propagation or SAT/SMT).
- **Score.** Gluability = (# covers with successful glue) / (total covers), with *repair cost* (minimal edits to achieve gluing). High-quality narratives have high gluability and low repair cost; incoherent or adversarial texts score poorly.

10. HoTT Identity for Linguistic Sameness

Linguistic "sameness" is intensional: we care about *how* two texts align, not merely *that* they do. HoTT models this with paths and higher paths.

- **Paths (paraphrase chains).**
  p : s1 =_S s2 witnesses that sentence s1 and s2 are the same in meaning.
  Basic laws: refl_{s} : s =_S s, inverse p^{-1} : s2 =_S s1, and concatenation p \cdot q : s1 =_S s3.
- **Higher paths (reasons-about-reasons).**
  \alpha : p \simeq q witnesses that two paraphrase routes p,q are themselves equivalent (e.g., two different proofs of synonymy).
- **Functoriality of sameness.**
  For any context/function f, ap_f(p) : f(s1) = f(s2). Meaning: substituting a paraphrase inside a larger construction preserves meaning.
- **Metric: Paraphrase Homotopy Index (PHI).**
  Build a paraphrase graph per proposition-type S. Let cycles modulo inverses count as nontrivial 2-cells.
  PHI = (# nontrivial homotopy classes of cycles) / (total cycles) (with a small-cycle cap). High PHI flags domains rich in paraphrastic structure rather than mere lexical overlap.

11. Transport, Funext, and Univalence in Concepts

11.1 Transport: substitution in context

Given p : a =_A b and a dependent family P : A -> Type, we can *carry data along sameness*:

transport^P(p, x) : P(b) , transport^P(refl_a, x) = x .

Example: with A = discourse entities and P = "truth-value of the story using this entity," replacing "she" by "Alice" via a proved coreference path preserves downstream truth.

Metric: Transport-invariance. Probe a model with paired inputs differing only by justified substitutions; require invariant predictions after transport.

11.2 Function extensionality: contextwise equality ⇒ equality

funext : (forall c:Ctx, M(c,s1) =_Val M(c,s2)) -> (M(-,s1) = M(-,s2)) .

Reading: if two expressions agree in every test context, they are equal *as functions of context*. Test by large, typed context batteries.

11.3 Univalence: concept collapse policy

We adopt univalence as an operational policy:

ua : Equiv(A,B) \simeq (A =_U B) .

A pair of concepts A,B may be *collapsed as equal* in the community's universe U only when we learn an equivalence with witnesses (round-trips homotopic to identities).

Metric: ua-witness test. Learn e : A -> B, e^{-1} : B -> A and demand

e^{-1} \circ e \simeq id_A , e \circ e^{-1} \simeq id_B .

Only then merge labels. Violations localize premature concept conflations.

12. Monads for Modality and Presupposition

Modal verbs, belief, and presuppositions act like *effects*.

- **Typing.** Meanings are Kleisli arrows A -> T B rather than plain A -> B, with monad T encoding the effect (e.g., possible-worlds, discourse context, presupposition store).
- **Composition.** Use bind:

f : A -> T B , g : B -> T C

a : A

in (f >>= g) a

Effects thread through the pipeline (e.g., common ground updates).

- **Laws.** Left/right identity and associativity of >>= constrain projection behavior (e.g., presupposition projection).

Metric: Effect-leak audit. Remove or rearrange clauses predicted to be *purely* intensional and check that truth-conditional outcomes elsewhere do not spuriously change; report leak rate vs. monadic predictions.

13. Polysemy as \infty-Groupoid Geometry

A polysemous word has interconnected, non-identical senses.

- **Model.** Senses are objects; sense-shifts are paths; reconciliations between shifts are 2-cells, etc. This captures *relatedness-with-distinction*.
- **Curvature intuition.** If many independent 2-cells relate two senses, the "space of explanations" is rich (harder disambiguation).

Metric: Sense-nerve curvature. Construct the nerve of senses and inter-sense transformations; estimate curvature via minimal 2-cycle density or discrete Ricci-style scores. Correlate with WSD difficulty and cross-domain transfer.

14. The CT/HoTT-Linguo Compliance Battery

We standardize a small, falsifiable test suite.

- **PHI (paraphrase homotopy):** cycle homotopy classes in paraphrase graphs.
- **ua-witness:** demand round-trip witnesses before concept collapse.
- **PCS (pullback coref):** fraction of gold links recovered by pullback selection.
- **BC-defect (scope):** preference margin for BC-permitted vs. blocked readings.
- **Gluability (discourse):** success rate of constructive sheaf gluing over covers.
- **Yoneda-probes:** typed context similarity as Nat(Hom(A,-),F).
- **Functor-consistency:** invariance under benign syntactic transforms.
- **Effect-leak audit:** rate of unintended side effects under composition.
- **Sense-curvature:** difficulty-predictive curvature of sense nerves.

- **Transport-invariance:** robustness under justified substitutions.
- **Curry-parity:** equality of readings under Hom(X \times A,B) \cong Hom(X,B^A).

Aggregation. Report each score per type/phenomenon; publish a macro-average and variance. Require seeds, templates, and transforms to be versioned for reproducibility.

15. Mini Case Studies

15.1 Scope and BC-defect

“Every student read some paper.”

R1 = \Pi_{s:Student} \Sigma_{p:Paper} Read(s,p)

R2 = \Sigma_{p:Paper} \Pi_{s:Student} Read(s,p)

Associate a square whose cartesianness licenses R1 <-> R2. Models should prefer the BC-permitted reading; BCdefect = margin(cartesian) - margin(noncartesian).

15.2 Coreference as pullback

“Alice met Beth. She smiled.”
Features: gender=fem, number=sg, recency, subjecthood.
Build f : Mentions -> C (features), g : 1 -> C (pronoun needs). Select P with f \circ p1 = g \circ p2. If m : 1 -> P picks Alice, PCS counts a hit; inconsistencies diagnose feature extractors.

15.3 Paraphrase lattice and PHI

s1: “The committee approved the plan.”
s2: “The plan was approved by the committee.”
s3: “Approval of the plan came from the committee.”
Paths p12, p23, p13; if p13 is not homotopic to p12 \cdot p23, we record a nontrivial 2-cell. PHI grows with such witnesses, marking rich paraphrase structure.

16. Related Work and Differentiation

- **Compositional formalisms.** Categorial/typed-lambda traditions model grammar–semantics composition; we align with them but *center universal properties* and testable functor/naturality constraints.
- **Monads in semantics.** Prior work treats modality/presupposition with effects; we add *effect-leak audits* as explicit behavioral checks.
- **Distributional approaches.** Usage-based vectors capture context; our **Yoneda-probes** turn "meaning-is-use" into *typed, falsifiable profiles*.
- **Sheaves, category-theoretic NLP, and higher structure.** Earlier sheaf/categorical models address discourse and glue; our novelty is integrating **HoTT paths/higher paths**, **univalence as a collapse policy**, and a compact **compliance battery** (PHI, PCS, BC-defect, Gluability, etc.) to move from metaphor to **measurement**.

17. Limitations, Risks, and Ethics

- **Idealization vs. data.** Real corpora are noisy; many "benign" transforms aren't strictly neutral. We therefore report *isomorphism up to specified tolerances*.
- **Coverage.** Our types are coarse; selectional restrictions and world knowledge may require richer categories or graded typing.
- **Metric gaming.** Benchmarks can be overfit. We mitigate via hidden templates, type-holdouts, and ablations.
- **Bias and fairness.** Coreference and concept collapse (ua) can encode social biases. Require audits across demographics and domains before deployment.
- **Interpretability tradeoffs.** Formal clarity can mask modeling uncertainty. All scores include confidence intervals and error analyses.

18. Outlook and Future Work

- **Benchmark release.** Open templates for PHI, PCS, BC-defect, Gluability, Yoneda-probes, with seed lists and deterministic transforms.
- **Richer types.** Add event calculus, time, and modality stacks; explore enriched categories for graded semantics.
- **Higher homotopy.** Go beyond 2-cells: quantify 3+-cell structure in paraphrase/sense complexes.
- **CT/HoTT-native models.** Train encoders that *enforce* functoriality, naturality, and monadic typing by construction, not as after-the-fact tests.
- **Cognitive hooks.** Use transport-invariance and funext tests as psycholinguistic probes: do humans and models share the same invariances?

19. References

Mac Lane, S. (1998). *Categories for the Working Mathematician* (2nd ed.). Springer. SpringerLink

Awodey, S. (2010). *Category Theory* (2nd ed.). Oxford University Press. Oxford University Press+1

Leinster, T. (2014). *Basic Category Theory*. Cambridge University Press. (Open-access ed. on arXiv.) arXiv

The Univalent Foundations Program (2013). *Homotopy Type Theory: Univalent Foundations of Mathematics*. IAS. arXiv

Lambek, J., & Scott, P. J. (1986). *Introduction to Higher-Order Categorical Logic*. Cambridge University Press. GitHub

Jacobs, B. (1999). *Categorical Logic and Type Theory*. Elsevier. People at MPI-SWS

Montague, R. (1973). The proper treatment of quantification in ordinary English. In *Approaches to Natural Language* (pp. 221–242). Reidel. Computer Science Department

Barker, C., & Shan, C.-c. (2014). *Continuations and Natural Language*. Oxford University Press. Oxford University Press

Shan, C.-c. (2002). Monads for natural language semantics. *ESSLLI Student Session*; arXiv:cs/0205026. arXiv

de Groote, P. (2001). Type raising, continuations, and classical logic. *Proceedings of the 13th Amsterdam Colloquium*. Loria Members

Moggi, E. (1991). Notions of computation and monads. *Information and Computation, 93*(1), 55–92. ScienceDirect

Plotkin, G., & Power, J. (2002). *Notions of Computation Determine Monads*. (Tech report/paper.) Informatics Homepages

Coecke, B., Sadrzadeh, M., & Clark, S. (2010). Mathematical foundations for a compositional distributional model of meaning. *Journal of Logic and Computation*. (Preprint: arXiv:1003.4394.) arXiv

Mac Lane, S., & Moerdijk, I. (1992). *Sheaves in Geometry and Logic: A First Introduction to Topos Theory*. Springer. SpringerLink

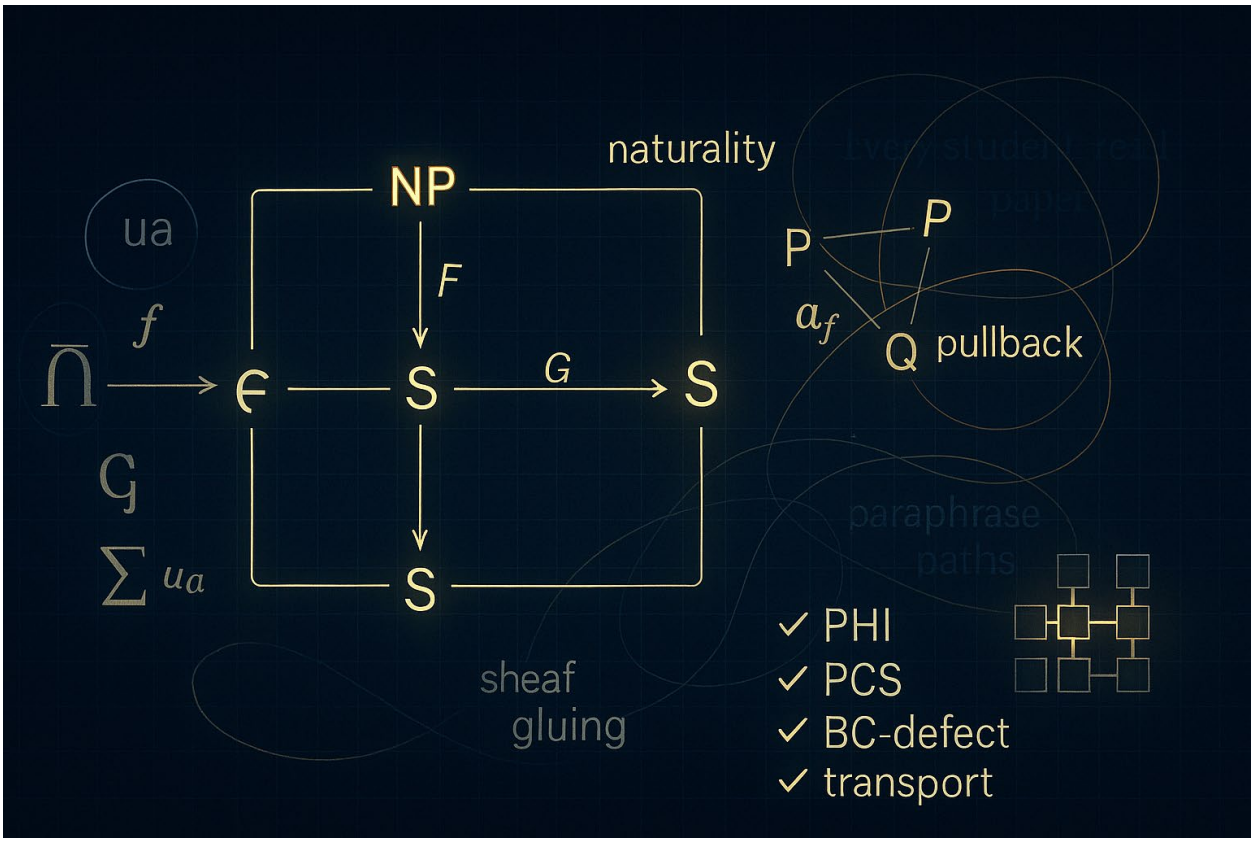
Abramsky, S., & Sadrzadeh, M. (2014). Semantic unification: A sheaf theoretic approach to natural language. In *CSL 2014* (Springer) & arXiv:1403.3351. arXiv+1

Prosorov, O. (2005). Sheaf-theoretic formal semantics (extended abstract). PDMI

Johnstone, P. (2013). Geometric morphisms of realizability toposes (on Beck–Chevalley contexts). *Theory and Applications of Categories, 28*, 235–268. tac.mta.ca

Jacobs, B. (2012). New directions in categorical logic, for classical, probabilistic and quantum logic. arXiv:1205.3940. arXiv

Youvan, D. C. (2025). *Achieving Artificial Intelligence (youvan.ai1) on a PC with Category Theory & HoTT*. ResearchGate preprint. DOI: 10.13140/RG.2.2.36085.03047. ResearchGate


naturality
NP
ua
F
f
G
S
S
S
P
P
a_f
Q
pullback
paraphrase
paths
sheaf
gluing
✓ PHI
✓ PCS
✓ BC-defect
✓ transport